对人类知识和学术的结构性和系统性分析

作者 顾立帆

（未经与我作认真和正式的确认，请不要采用我文章里的任何观点和想法，以免产生任何误解或曲解）

人类的知识和学术，经历了从经验、到数学、到科学的发展。其间不断受到蒙昧主义、诡辩等等的困扰和误导。亚里斯多德谈到了第一性，并建立逻辑学试图清除诡辩。

但二十世纪的新研究证明了逻辑学不但不具有科学和智力上的第一性和完整性，逻辑学本身也有问题，并无法真正识别和解决很多重要的诡辩问题。事实上，人类主要的学术分支：数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式的错误。而计算机思维模式的问题更加严重。

计算机无法处理高阶逻辑，而只能使用过度简化的逻辑，所以不能处理诡辩，也不具有科学和智力上的第一性和完整性。而诡辩会在司法、医疗、教育、研究等等方面造成严重问题。从小习惯计算机过度简化的逻辑和流程的学生，会缺乏第一性和完整性的教育训练、缺乏系统性和抽象性的思考能力等等，其智力发展会受到损害。

至少从 leukotomy 开始，脑和智力研究就陷入错误无法纠正的困境。在人工智能兴起的今天，更出现了混淆数学和科学，混淆经验和科学，混淆经验测试和科学实验，把亚里斯多德思维模式当作正确的思维模式等等错误，在人工智能的哲学基础和测试理论上都有严重问题 [1]。拒绝别人用科学实验对经验结论作证伪，也使诡辩在学术界盛行。诡辩会混淆是非对错，使学术研究失去正确的判断力、误导未来发展。而智能技术的发展，也对 Total Quality Management 形成了新的挑战。

本文从分析亚里斯多德思维模式的错误开始，分析 leukotomy 的本质问题和对基因工程的深刻教训，研究人工智能在哲学基础和测试理论上的问题，澄清科学与经验和数学的差别，在此基础上用科学的方法对人类知识和学术作结构性和系统性的分析，以此澄清学术上的众多错误，并导出对语言处理、思维模式、科学和教育发展的新理论。

科学研究基于平等讨论和公正实验。本人设计了一个关于教育、科技、和文明发展的讨论；也设计了严格控制条件、渐进地测试人工智能的科学实验，并对智力作结构性和系统性的分析，对第一性和完整性作复杂性的分析，形成智力研究的科学基础。

一、亚里斯多德思维模式在学术研究上的错误

有研究指出亚里斯多德建议亚力山大: to be "a leader to the Greeks and a despot to the barbarians, to look after the former as after friends and relatives, and to deal with the latter as with beasts or plants"。所以，亚里斯多德思维模式明显不具有完整性。

在学术研究上，现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重问题和错误。

物理学和化学，本来就是建立在对亚里斯多德思维的批判上。亚里斯多德思维模式凭经验就草率推断。而科学要求严格控制条件的证伪实验作基础。

达尔文的演化理论，说明了物种是动态发展的，亚里斯多德的静态语言方法和思维模式忽视了生命的本质。

亚里斯多德基于特征的生物分类法，是经验性的主观方法。这种主观方法会掩盖分类法中的歧义。现代生物分类法试图引入基因等客观因素，却因此暴露了生物分类法中的歧义是本质性的，无法用逻辑来消除。

基因研究揭示了遗传和变异的复杂本质，但也给生物分类带来进一步危机。基于基因的生物分类虽然比基于特征的生物分类更准确、客观，但基因的渐变、交叉混杂、以及和其他因素的交互作用，也使正确划分物种界线变得困难，甚至在顶层类别划分上都充满争议，其歧义无法用逻辑消除。

幸运的是，生物分类对很多方面的生物研究只具参考价值。很多研究并不真正需要依赖生物分类的精确定义。所以，生物学的很多部分，能够作为严谨的科学。

哥德尔事实上证明了逻辑学不具有数学上的第一性和完整性，即使是数学也不符合亚里斯多德的思维模式，其发展需要更高的智力因素。甚至连康德关于人和思维的哲学，都已经突破了亚里斯多德的思维模式。

所以，近年有人试图把亚里斯多德当成学术上的正统，抬高第谷而无视第谷理论体系的错误 [2]，正在把学术引入歧途，特别是在脑和智力研究上。

下面从 leukotomy 开始，分析脑和智力研究上，亚里斯多德思维模式在哲学基础和验证方法上的问题。本文不讨论亚里斯多德思维模式在学术研究以外是否有用。

二、**Leukotomy** 错误的本质及其对基因工程的教训

Moniz 获得诺贝尔医学奖的理由是: "for his discovery of the therapeutic value of leucotomy in certain psychoses."。所以 leukotomy 的学术错误，应该从 leukotomy 到底能治疗什么精神疾病、会造成什么严重的不良后果、获奖前有没有严格地验证来分析。

事实上，leucotomy 没有真正治愈任何疾病。它是一种不针对具体疾病来具体分析的盲目手术：用切断重要脑组织连接来消除大量的脑活动。因为没有对准具体的疾病，把很多没有疾病的人也错当手术对象。因为盲目切断脑连接，损害了重要的脑功能。所以，leucotomy 不但本身是一种错误的手术，在学术上也没有多少价值。后来较成功的精神疾病治疗方法，需要针对具体的疾病来治疗，不能沿用 leucotomy 不区分清楚对象是否真有疾病、或者真正的疾病是什么而盲目切除的方法。

Leukotomy 切断的脑连接，并不是多余或无足轻重的，而会造成严重问题。问题没有引起足够重视的原因包括：

1）选择手术对象时没有正确区分 personalities 和 mental diseases。

2）给 leukotomy 授与诺贝尔奖时，没有正确地检验 leukotomy 的后果，对提出警告者也没有重视。

3）手术对 personality 和 intelligence 有严重损伤。连大规模实施者 Walter Freeman 都在一定程度上承认。由于缺乏对 personality 和 intelligence 的理解，事后的统计无法反映其损伤 [3]。

无法正确理解 personality 和 intelligence，无法正确区分 personalities 和 mental diseases，无法正确检测对 personality 和 intelligence 的损伤，是由于缺乏正确描述 personality 和 intelligence 的语言，缺乏这种前瞻性的新领域研究所需要的哲学性思考和分析，而造成的无知。

由于目前对 personality 和 intelligence 仍缺乏研究，仍无法正确区分 personalities 和 mental diseases，无法正确检测对 personality 和 intelligence 的损伤，所以对人类的基因改造和基因控制有巨大危险。

Leukotomy 获得诺贝尔奖之前的研究，没有作严格控制条件的科学实验和检测，仅仅凭经验性的手术尝试作出主观的评估。这种主观评估跟后来发现的灾难性后果明显不符。

而缺乏严格控制实验条件的主观评估，也正是 Turing Test 的问题之一。Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，也是主观评估。这个 superhuman 是针对 generic 的 human，既缺乏严格控制条件的实验依据，又不接受别人设计的证伪实验。

Leukotomy 的诺贝尔奖有错不纠，和 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称拒绝严格控制条件的实验证伪，显示 peer review 有严重问题。Peer review 也是主观评估 [4]。

所以，leukotomy 的问题，实际上就是亚里斯多德思维模式的问题。这种问题已经危及脑和智力研究及相关领域。以下第三至五节，讨论跟人工智能有

关的各种问题。

三、AlphaGo Zero 的 Superhuman 宣称不是科学结论

Nature 杂志上的 AlphaGo Zero 论文，宣称其取得 superhuman 的 performance。但 superhuman 的概念是针对 generic 的 human。该论文并没有提供任何能证明 superhuman 的证据。

AlphaGo Zero 击败 AlphaGo Master 并不是 superhuman 的证据，因为两者都存在人工智能的某些类似缺陷；电脑围棋在一些 game 中胜过对计算机和人工智能的本质问题尚缺乏了解的棋手，也不是 generic 的 superhuman 的证据。这种 game 只是经验测试，不是科学实验。科学实验需要严格控制实验条件才能推导出科学结论。

AlphaGo Zero 用算法或数学方法来快速模拟对弈、积累经验、提高棋力，而用经验方法来测试成效。这些模拟对局的数量虽然很大，但跟围棋所有可能走法的数量相比，却微不足道，远远不够覆盖大部分的可能性。所以 AlphaGo Zero 在绝大部分的情况下，并不知道正确（即最好）的走法，只是靠特征提取和概率性估算来猜测胜负的可能性。

传统人类棋手的棋力，基于经验性的智力，所以在绝大部分情况下也不知道正确走法。水平高的棋手知道比较好的走法，但不知道最好的走法。AlphaGo Zero 经历的模拟对局数，远远超过个人下过的对局数。AlphaGo Zero 能总结出的棋力也远远超过人。即使人能抓住 AlphaGo Zero 的错着，但人的错着更多、错误更严重。到最后往往还会是 AlphaGo Zero 赢。要抓住 AlphaGo Zero 的错着，并把优势保持到最后，是很难的。

所以，电脑围棋目前在棋赛中击败人类棋手，实际上是数学对经验的胜利。但人的智力并不仅仅限于数学和经验。我设计了 4 维实验空间的科学实验来验证电脑围棋存在大概率的问题，让人类棋手有机会发挥他们在抽象思维和其他方面的优势 [5]。

很多人把深度学习的特征提取，等同于抽象思维。我设计的实验，可以验证两者之间的区别；可以验证对所提取特征的匹配应用，其实只是概率性猜测，并不可靠。

目前对电脑围棋研究的一个主要问题，是无法正确判断其走法的对错。即使怀疑电脑围棋某些走法是错着，但如果人类棋手最后输了，就难以判断电脑围棋的那些走法是真正的错着、还是另有胜机的妙着。电脑围棋的特征提取和概率性估算也说明不了问题，因为它也是猜的，并不知道真正的正确走法。

不知道对错就无法确定电脑围棋的真正问题，当然更难以改进。我设计的实验空间，则有方法可以在公平对等的情况下，暴露电脑围棋一些大概率的真正错着和问题，帮助人类棋手深入研究电脑围棋的问题，帮助学术研究者研究人工智能的系统性问题。

科学是为了弥补数学的不足而发展起来的。一般情况下，科学方法优于数学方法，而数学方法优于经验方法。所以，我建立了智力的结构性分析理论，来区分科学、数学、经验等等的性质特征，以此设计验证 AlphaGo Zero 和其他人工智能问题的科学实验。无人汽车和波音飞机的致命事故，说明我的结构性分析理论和测试方法非常必要。

有些人类棋手已经发现电脑围棋的一个缺点，就是其特征提取和概率性估算实际上没法精确计算一些对杀的后果。而人能精确计算一些对杀。有些对杀对输赢影响极大，一旦算错，以后就没法扳回了。当然电脑围棋计算能力很强，用特征提取和概率性估算已经能找出很多对杀方法。出现这种估算错误的可能性有多大，仍有待研究。而我设计的实验空间，是针对系统性的大概率问题，不是那种零星的估算错误。

我设计的实验，实验结果和结果分析会全部公开。别人可以对实验过程和结果作分析、推敲。这样才是真正的科学实验方法。如果人类棋手在我设计的公正实验条件下也能胜过 AlphaGo 系列，意味着 AlphaGo 系列所依赖的人工智能技术有严重的缺陷 [6]。这对脑和智力的研究、以及人工智能的健康发展都有很大价值。

我设计的实验，基于 Turing Machine 和人工智能的局限性，不需要 AlphaGo 系列论文和棋谱中的信息。所以可以验证人工智能和计算机层次的局限性；所以即使别人向我提供了错误的系统信息，我的实验仍能找出真正的问题。但出于对科学和公众负责，论文作者应该公布跟论文有关的全面准确

的实验结果：包括 AlphaGo Zero 对 AlphaGo Master，和最后版或稳定版的 AlphaGo Zero 对 AlphaGo Zero 的更多棋谱 [7]，等等。

除去不真实的宣传，AlphaGo Zero 并不比二十年前的 Deep Blue 更重要。AlphaGo Zero 所依赖的主要技术新进展，也来自其他方的研究。和 Deep Blue 不同，Demis Hassabis 经常毫无依据地宣传他们的人工智能是以解决所有问题为目标 [8]。所以需要对 AlphaGo Zero 的 superhuman 宣称设计严格控制条件的科学证伪实验，纠正教育和宣传中的误导。科学研究，需要不断接受责疑和改进，在证伪中进步。

本人无意贬低电脑围棋的进展。相反是鼓励电脑围棋的研究，并将其作为基础智力研究的实验工具。早在 Deepmind 成立之前，我就分析认为电脑围棋可以击败人类顶级棋手，因为围棋是在很小空间里的高度规则化游戏 [9]。这已被证实。但这并不意味电脑围棋有超人能力或能解决一切问题的潜力。

基于数据的人工智能或统计方法，无法解析未来的模式演变。在模式演变不稳定的情况下，会产生严重误导。有些以前几乎不出现的问题，模式演变后，甚至可能成为大概率事件。对智力作基础理论分析，才能了解智力的模式演变是否稳定。

四、自动驾驶的定义和测试问题

本人以前写过:“Go gaming is strictly defined within a very small space. Industrial automations are typically designed in well controlled environments, but not strictly defined. Car driving is regulated, but the environment is not well controlled”。

自动驾驶汽车中的问题，远比电脑围棋复杂。如果在电脑围棋中，不稳定的模式演变，都会造成大概率的严重问题，采用类似人工智能技术的自动驾驶就更需要谨慎了。所以需要验证最强的电脑围棋（如 AlphaGo Zero）在不稳定的模式演变下的问题。

无人汽车不加限制说明，就是 SAE level 5 自动驾驶，而这是不可能的。这可以通过对智力原理的分析和对人工智能所使用的技术设计实验来验证。而 SAE level 4 的自动驾驶，在不同的驾驶区域，实际上有不同的技术要求。所以 SAE level 4 的定义，会隐含很多问题。即使在同一区域，未来不稳定的模式演变，也可能使已通过 SAE level 4 测试的自动驾驶系统，出现大概率严重问题。

对无人汽车的模拟测试和路试，无法覆盖或演绎出大部分的情况，特别是对未来不稳定的模式演变。即使在封闭的 19X19 围棋小空间，AlphaGo Zero 都无法覆盖绝大部分情况。这可以设计严格控制条件的实验来验证。无人汽车面对更加复杂的环境，其人工智能中的经验和数学方法，也会造成更严重的问题。所以更需要科学方法的检验。

科学通过抽象和对抽象原理的证伪来间接 address 所有情况。科学实验通过对实验条件的严格控制才能验证和改进原理。所以无人汽车的模拟测试和路试不是科学实验。

另外，无人驾驶汽车的测试，完全不同于其他汽车。汽车中的动力（燃料或电动）、机械、电子等设计，早就具备坚实的科学基础。而无人驾驶的人工智能，没有科学基础，也未经严格控制条件的科学实验验证，危险性很大。

现在使用的是经验测试，无法解析未来模式的演变。而普通人可能需要很多年跟无人驾驶汽车的大规模互动，才能接触到其临界点，引发重大的模式演变。所以需要对智力的基础理论作前瞻研究，设计严格控制条件的科学实验来验证和证伪这些技术。

本人不直接研究自动驾驶。而主张对智力作系统性、结构性、和复杂性的基础分析，对电脑围棋等等人工智能技术设计严格控制条件的证伪实验，来研究不稳定的模式演变，为智力研究建立科学基础。能这样在不引发大规模严重后果的情况下，用基础研究的方法，为自动驾驶技术作前瞻性研究，其实是一种幸运。

本人赞成研究电脑围棋，也赞成研究自动驾驶。只是不赞成过度炒作无人汽车，不赞成把未经严谨的证伪科学实验检测的人工智能技术用于商业化的自动驾驶。

下面分析人工神经网络和预测术的本质，分析基于经验和数学的人工智能和科学的本质不同。

五、预测术、人工神经网络、和计算机技术的局限性

人工神经网络实际上是对已有信息的改写和预测。

人类文明的早期，预测术曾被普遍看重。中国古代有用于预测术的易经：《连山》据说是夏朝的易经，《归藏》据说是商朝的易经，以及周朝的《周易》。

商代夏，周代商，模式演变大致是稳定的。旧易经不对，就换新易经。到春秋战国的混乱取代周朝的实际统治时、和秦国结束战国混乱时，模式演变是不稳定的。没有人再能写出流行的新易经。预测术虽仍继续存在，但已远不如夏、商、周时那样重要了。

《道德经》中的“道可道，非常道。名可名，非常名。”实际上是对预测术的否定。即使是现代的科学之道，也需要不断发展。

到欧洲文艺复兴后期，由于天文学的发展，占星预测术又开始流行。

第谷的天文理论体系是错的，却得到了很多资助。开普勒的天文理论体系是对的，却死于贫困。开普勒的天文学研究，在其生前并不受人重视，不得不靠占星预测术来争取收入，维持他的天文学研究。

占星预测术被认为是实用的。而基础天文学研究对一般人没什么用。但天文学能够逐渐发展成为非常严谨的学术，而占星预测术到了今天仍然仅仅是预测术而已。

到二十世纪前期，著名心理学家荣格借助星座分析人的性格。荣格的星座分析，已经和开普勒的占星预测，有了很大的不同。但即使荣格在他的星座分析中引入了很多哲学性的思考，但他的星座学仍达不到哲学完整性，当然更达不到科学的标准。

并不是说预测术就不能研究。但需要区分预测术和科学的区别。也不应该用预测术的研究，来排斥对智力基础理论的科学研究，即使基础科学研究在早期显得没什么用。

预测术是基于经验或数学的。哥白尼、第谷、开普勒时期的天文学，也是基于数学的，还不是严格意义上的科学。开普勒把哥白尼的行星轨道从圆形改成椭圆形，但现代科学发现行星轨道并不是椭圆的 [10]。确定更精确的轨道，需要对力的原理进行分析。所以，哥白尼和开普勒都不是严格意义上的科学家。

爱因斯坦追随的是伽利略的哲学、和他对原理的研究方法。这不同于牛顿的哲学。所以科学的创建者是伽利略，并不是哥白尼、开普勒、或牛顿。

科学基于对抽象层次的原理分析，而不是对现象层次的预测作验证。现象层次的预测，缺乏普适性，对模式演变不稳定的情况难以分析和验证。需要用严格控制条件的实验对抽象的原理作证伪。只有通过对控制条件的抽象才能推导出科学结论。而证伪也包括寻找控制条件下的新反例。所以，科学和经验及数学都有本质不同，也不是预测术。

而可以抽象成 Turing Machine 及其各种变种的计算机技术，也有本质的局限性：既不具有高层次的完整性，也不具有科学上的证伪能力。Universal approximation theorem 的名称，也很容易误导。定理中所能趋近的函数，是有严格限制条件的。并不是所有函数都能无限趋近。基于传感器的人工神经网络也是经验性的，不具有完整性。

近年有人根据人工智能的一些成功经验，炒作 Technological Singularity，实际上是对科学哲学的无知。人工智能最大的危险，不是其太强大可能控制或消灭人类，而是大规模应用不成熟的人工智能可能带来巨大灾难。

六、脑和智力研究需要新的科学哲学

理解了科学和经验及数学的本质不同后，就可以分析目前脑和智力研究上的各种问题，澄清哲学和思维模式上的各种错误。

1949 年诺贝尔生理医学奖 leukotomy 存在严重的检测错误，并造成灾难性误导，但至今无法纠正。科学研究上有错不纠，会蔓延成脑和智力研究上的普遍问题。

Turing Test 和主流教科书 AI: A Modern Approach 有哲学基础和测试理论上的严重问题，把 Wind Tunnel approach 应用于人工智能已造成严重后果；权威学术杂志 Nature 上 AlphaGo Zero 论文中的 superhuman 宣称，破坏了学术语义和科学实验的严谨；Technological Singularity、In Math We Trust、和“人类正在扮演上帝的角色”的观点 [11]，甚至出现在专业研究人士中间；而把脑研究比作登月项目的官方观点，更是惊人的无知 [12]。

种种问题说明：现实中真正奉行的研究思维和方法有严重隐患。科学哲学在 Karl Popper 以后，基本停滞不前，不但导致脑和智力研究在理论上误入歧途，也使科学研究在整体结构上越来越难以突破，甚至错误泛滥。

很多耗资巨大的实验结束后，既证明不了什么本质，也否定不了什么本质，使新的科学理论无法成熟。而对经验性的人工智能技术，拒绝别人的证伪实验，也使诡辩在学术界更加流行。学术语义和科学实验的严谨一旦被打破，教育、学术研究、医疗、和司法等等许多方面，都会出现全局性严重后果，危及未来的发展 [13]。

忽视、排斥科学哲学的人，很可能用一种错误的哲学来看待科学，比如经验哲学、实用主义、Occam's Razor、逻辑实证主义，等等。

Occam's Razor 仅仅是一种经验的哲学方法，不但不是科学，也不是严谨的科学哲学。所以从几种无法证伪的理论中，选择最简单的理论，并不是科学结论。对 Occam's Razor 在学术研究中的广泛使用，需要科学哲学的新研究来澄清其隐含限制、限定其使用。

科学理论的简单和复杂，甚至不一定存在数学上的全序关系。所以，几种暂时无法证伪的科学理论中，也不一定存在最简单的理论。即使存在数学上的全序关系，Occam's Razor 所剃掉的是不是真正多余的部分，有可能需要很长时间才能发现，不能草率下结论。像 Leukotomy 切断的脑连接，并不是多余或无足轻重的。就是一个深刻教训。

科学的基础是证伪，不是经验。不但亚里斯多德思维模式在学术研究上是错误的，连逻辑实证主义也有问题。逻辑判断使用的语言和经验证据之间可能有很大的差别，经验证据往往不能真正支持逻辑判断。而科学的原理分析，可以发现两者之间的差别。基于科学原理分析的证伪实验，可以检验这种区别。

所以，对人类在历史长河中形成的林林总总的经验、知识、学说，及研究领域中的很多流行观点等等，需要从科学哲学的新高度去分析、理解、判断其对错和长远价值。对人工智能的实用性、可靠性、安全性，也需要用严格控制条件的科学实验来证伪。

以下两节介绍本人对人类知识和学术的结构性和系统性的分析，作为智力研究的科学哲学基础。而把智力研究提升到科学层次，则需要严谨的科学实验基础。Gu Test: A Progressive Measurement Of Generic Artificial Intelligence 一文介绍的是本人的实验设想。

七、对人类知识和学术的结构性分析

语言是人类重要的智力特征。其他已知生物，都不具有复杂的语言能力。

在语言基础上，人类发展出两类重要的学术：数学和哲学。数学向精确化发展，哲学向完整性发展。

数学是一门非常古老的学科。而科学则是近几百年中为了纠正数学的不足才发展起来的。科学起源于哲学。所以，严格意义上的哲学和科学，具有哲学完整性。科学也不断引入数学和精确性，但科学的本质特点是严格控制条件的证伪实验基础。数学达不到科学标准，也不具有哲学完整性 [14]。

达尔文的演化理论，可以通过科学实验作验证，所以是科学理论。而他的物种起源理论和宇宙大爆炸理论虽从科学理论推导而来，却仅仅是历史哲学层次的准科学。虽然可以用到很多用科学方法收集的证据，但历史结论本身无法用严格意义上控制条件的科学实验来证伪。所以达不到科学理论的标准 [15]。

生物分类法的本质歧义，无法用逻辑来消除。所以达不到哲学完整性 [16]，仅仅是语言层次的知识，只能作粗糙的参考。最多算准哲学。很多重要的生物学理论并不依赖生物分类的精确定义，所以可以作为严谨的科学。

研究智力的基础理论，需要精确理解和区分经验、语言、哲学、数学、科学等等不同层次、特征、和标准的理论和方法，研究其各自发展的基本原理。

一般情况下，科学方法优于数学方法，而数学方法优于经验方法。但人类仍需要经验和数学的方法。因为在很多情况下，可以运用经验方法，但不存在有效的数学方法；或者存在有效的数学方法，却不存在有效的科学方法。也就是说经验方法的适用范围比数学方法更广，而数学方法的适用范围比科学更广。

现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重问题和错误。现在需要跨学科对亚里斯多德思维模式的问题和错误，作科学哲学上的进一步剖析、抽象、和总结，作为智力研究的哲学基础。

开普勒的占星预测术、荣格的星座分析学、人工神经网络，也达不到哲学完整性，当然更不是科学。而计算机不具有人的 commonsense，无法处理高阶逻辑，只能使用过度简化的逻辑。所以，计算机思维模式的问题，比亚里斯多德思维模式更加严重。

对理论标准，作经验层次、语言层次、哲学层次、数学层次、科学层次上的严格细分，是为了建立严谨的科学哲学和科学体系，和精确地衡量人类智力的不同范畴和方法、及所作结论的不同性质。

下面对数学和科学的历史演变及其重要临界点作分析，研究人类主要学术体系第一性和系统性的发展、演变规律。

八、数学和科学的演变及其重要临界点

从毕达哥拉斯到伽利略，是人类文明走上可持续发展的关键。

毕达哥拉斯的抽象化思维，创立了数学。他对算术和数的定义和体系，按现代抽象代数的概念也是成立的，但不适用于几何和其他一些系统。毕达哥拉斯算术系统本身在科学上是错的，但在数学上又是可以成立的、有价值的、可以继续发展的。所以数学不同于科学，也不能简单用数学来验证科学上的对错。

希帕索斯的发现，本来也是对数学和毕达哥拉斯体系的发展。但希帕索斯事件之后，毕达哥拉斯学派本身就没有太多成就了。其他群体则能继续他们的工作，后来居上。

欧几里德几何创建了重要而又可以完备化的数学理论体系，并带动了阿基米德等等一大批泛希腊化时代的学术研究者。

毕达哥拉斯和欧几里德两千年之后，有牛顿和莱布尼茨创建微积分，罗巴切夫斯基等人推翻欧几里德几何的唯一性，建立非欧几何 [17]。

朗道实际上把自己当作第四流的科学家：牛顿是第一流，爱因斯坦是第二流，创建量子物理的才俊们是第三流，和朗道同时代的美国物理学家理查德·費曼也是第三流。也许朗道是自谦，但他肯定无法和爱因斯坦相比。而俄罗斯历史上是出过门捷列夫的。

不知朗道该如何评价伽利略、达尔文。

按照我对学术成就的评价，创建重要学科、和重要学科中的理论体系，是第一流的工作；在重要学科中提出革命性理论体系，是第二流的工作；对主要理论体系的重要延伸或修补、或开创较小的学术分支，可列入第三流的成就。第一、第二流的学术成就非常罕见，第三流的工作已经非常了不起了。这个标准跟朗道的评价高度吻合。

物理学中，伽利略和牛顿的贡献，跟数学中的毕达哥拉斯和欧几里德相仿，属第一流。相对论和量子力学是重要学科的革命性理论体系，跟数学中的微积分和非欧几何相仿。所以，爱因斯坦的贡献是第二流的，而创建量子力学有许多人的贡献，他们的贡献只能算第三流的。量子电动力学只是小分支，所以理查德·費曼的工作也是第三流的。

牛顿和莱布尼茨在数学中的贡献是第二流的。但非欧几何的创建，由多人逐步完成，有许多人的贡献。罗巴切夫斯基和鲍耶对推翻欧几里德几何的唯一性最重要，他们和希帕索斯（发现无理数）、亚里斯多德（逻辑学）、帕斯卡（概率论）、哥德尔（哥德尔定理）相仿，仅略逊于牛顿和莱布尼茨。

拉格日朗（数学分析的成熟化）、图灵（图灵机作为计算模型）的贡献是第三流的；数学中的高斯、和物理学中的朗道，能不能挤进第三流，我不知道；

祖冲之的贡献，未必能超过高斯。但有观点认为《九章算术》首次引入负数 [18]。此说如果确实，《九章算术》的数学成就，也许并不比希帕索斯和亚里斯

多德逊色很多。这可能跟中国古代的阴阳玄学哲学性观念有关 [19]。

数学是非常古老的学科，现在看起来很重要的一些成就，跟数学历史长河中的很多大师比起来，不一定更重要。十七、十八世纪的数学，和二十世纪前半期的物理学类似，已经十分兴旺，但基础是更早的人奠定的。

达尔文跟毕达哥拉斯和伽利略相仿。孟德尔不如欧几里德和牛顿，因为生物学整体理论在不久之后还需要几次重要的综合才能形成。

波义耳和门捷列夫在化学中的作用，有点类似于物理学中的伽利略和牛顿。但门捷列夫和波义耳不但时间相隔较久，而且中间有很多人的重要贡献。所以，他们的学术贡献，也不如伽利略和牛顿。

这里的评价，全部按照学术贡献的大小。不是评估个人智力，也不是评价人的尊卑。而是评估其工作对人类整体智力发展的影响。人各有特长，个人智力也不一定都表现出来，很多情况下是不可比较的。而学术贡献，是已完成的工作，可以客观分析。

整理历史上学术贡献的大小，一方面是为了研究人类智力的演变和其中重要的临界点，另一方面也是研究能促进未来发展的各种因素。现在把智力研究转化成科学的关键，在于形成准确描述智力的语言概念、哲学基础、和数学形式。

如果物理学、化学、生物学、人工智能或现有数学等等，都不能提供这些概念、基础、和形式，就需要创建新的学科。能不能提供，可以通过证伪实验来验证。

科学和数学的本质和方法，有所不同。科学兴起虽然较迟，但能弥补数学思维方法的一些严重不足。事实上，物理科学比数学更依赖于哲学性思考。而生命科学的哲学性，比物理科学更重。开创智力新科学，比起生命科学也更具哲学上的挑战性。

从数学、到物理科学、到生命科学、到智力科学，随着哲学性思维比重的不断加大，其研究和教育也会越来越平民化，可以为普通人带来更多的机会。因此也需要探索智力科学时代平民化教育和研究的新方式。

新的哲学思维，也会带来思想上的新挑战。而希帕索斯和伽利略因为追求学术真相而受迫害，更应引起人们的警觉。相比之下，高斯虽是优秀的数学家，却缺乏追求学术真相来推动进步的求道精神，历史地位有限。

公正的研究环境、和免受学术迫害，是科学能够持续发展的前提。

九、第一性和完整性的结构性和复杂性研究

亚里斯多德虽然谈到了第一性的问题，但欧几里得几何公理体系显示了逻辑学并不是数学公理体系中的第一性。而哥德尔证明了逻辑学确实不具有数学上的第一性，当然也不可能具有科学和智力上的第一性。

而科学的建立和发展，说明了数学也不具有科学和智力上的第一性。科学需要苏格拉底方法、科学实验和其他途径获取第一性。

毕达哥拉斯试图建立一个具有完整性的数学体系，而希帕索斯显示了这个数学体系并不具有完整性。欧几里得几何公理体系似乎具有完整性，但非欧几何显示几何公理体系的完整性是局部性的，数学整体上并不具有完整性。而哥德尔更证明了任何数学体系都不可能具有数学上的完整性，当然更不可能具有科学和智力上的完整性。

对第一性和完整性作结构性和复杂性的进一步研究，可以分析各种智力体系的能力和局限性，帮助建立智力研究的科学基础、发展出新的科学哲学。

十、语言处理、语言发展、和科学发展

语言是人类知识的重要载体，也是哲学、数学、和科学的基础。

伽利略完成了对静止和运动的抽象，才能形成对力的正确理解，才能形成物理学中的科学语言。

根据对人类知识和学术的结构性和系统性分析，可以建立对语言处理、语言发展、和科学发展的新理论。

人类知识和学术的发展，依赖于语言的发展，特别是在重要的临界点上。

人工神经网络本质上是对已有信息的改写和预测，在低层次的自动翻译上可以取得很多进展。但这种预测并不是科学方法，无法发展科学的新思想，甚至也无法提供科学发展所依赖的语言发展。

计算机和人工智能技术，包括人工神经网络，无法正确处理高阶逻辑，只能使用过度简化的逻辑。所以，计算机思维模式不但无法识别处理诡辩和亚里斯多德思维模式等等问题，还会造成更严重的问题。还有 Chinese room 的问题，不能真正理解语言。

人类智力和人工智能、计算机有本质的不同。其各自的长处、短处也很不相同。所以，用给人类考试的试题来测试人工智能，参考价值很有限。实际上反而会掩盖人工智能的很多关键性问题，很容易引起误解。

现在测试语言理解的 dataset 如 SQuAD、CoQA、QuAC、和 GLUE 等等系列，可以检测较简单的阅读问答，但不能测试真正的理解，也不能测试用语言深入思考的能力。因为缺乏对语言智力的科学和结构性分析，这些 dataset 无法真正衡量人工智能和人类智力的差距，也不能在诡辩识别、高阶逻辑处理、Chinese room 等问题上对人工智能有多少提高。

NLVR^2（Natural Language for Visual Reasoning for Real）等 testset 虽然难度较高，但仍缺失对重要语言智力因素的测试。GLUE 也并不具有测试 generic 语言智力的能力。

事实上，所有这些 Testsets 都陷入亚里斯多德思维模式的陷阱，都不是严格意义上的科学测试方法。所以说，主流教科书《AI: A Modern Approach》在哲学基础和测试理论上的错误，覆盖了人工智能整个行业。

现代主要学术分支数学、物理学、化学、生物学，均已宣告亚里斯多德思维模式有严重错误。而计算机思维模式的问题更加严重。如果不能识别诡辩、亚里斯多德思维模式、计算机模式等等的问题，而错误地把人工智能用作 guideline 或 judgement，会误导司法、医疗、研究、和教育等等，损害智力的发展。

所以需要研究语言、哲学、数学、科学的发展，研究智力的基础理论，建立对语言处理的科学测试方法，验证 Chinese room 问题，检测对高阶逻辑的正确处理和局限性，检测识别诡辩、计算机思维模式、和亚里斯多德思维模式等等问题的能力。

近代和现代制造业的发展，是建立在科学进步的基础上。而计算机和信息技术的功能，可以抽象为数学的子集，所以其实是数学技术。数学不能取代科学。计算机和数字化产业虽然使数学获得了新的活力，但并不能取代以科学为基础的现代制造业及其技术。而近年基因技术的发展，实际上也给农业带来了新的机会。

The Third Wave 之类的商业畅销书，对人类学术和文明发展，缺乏结构性和系统性的理解。不但混淆了数学方法和科学方法的差别，而且过分简单化地划分第一、第二、第三产业。服务业的各种服务，差别很大。很多部分跟计算机和信息技术很不一样。把这些笼统地划为第三产业，暴露出缺乏哲学基础的流行语言的问题。

而计算机和信息技术所用的设计方法，特别是过度简单化的逻辑方法，也和科学研究方法有很大的不同。人工智能所用的数学和经验测试，更是不同于科学方法，也不能得出科学结论。在对 AlphaGo 系列和 self-driving car 的测试上，就充分暴露出来。科学需要更高的抽象程度对实验条件严格控制，以此才能推导出科学结论。

The Third Wave 一书不但模糊、混淆了计算机和信息技术所用的设计方法和科学方法的差别，而且把计算机和信息技术当作第三次浪潮的标志、不正确地引导出计算机和信息技术对制造业及其科学技术存在新、旧交替的错误观念，不但给美国制造业带来严重灾难，也给全世界科学哲学和智力基础科学的研究发展方向带来严重误导。

最近波音飞机的问题，更说明忽视制造业的后果。制造业的发展，需要建立在科学进步的基础上。计算机和数字化技术，不能取代科学，更不能解决一切问题。

人工智能，特别是人工神经网络之类的技术，有潜在的不稳定性。这也给 Total Quality Management 带来了新的挑战。

所以，商业畅销书不能取代科学哲学的研究。智能学科的进一步研究和发展，需要新的哲学基础。

物理学中有海森堡测不准原理。所以临近测量极限，存在很多争议，有些甚至是无法解决的。近年有很多物理实验耗资巨大，却没多少收获，给科学研究造成很多迷惘。

而智力现象跟人的日常生活和发展密切相关。只是目前的人工智能是经验技术，缺乏严谨的科学理论和实验基础。我的理论和实验设计，是要在实验空间中寻找远离测量极限的临界点，为智力研究建立渐进的、严格控制条件的实验基础和科学原理，为科学研究带来新的生机和动力。

十一、教育的变革

诺贝尔生理医学奖 Leukotomy 的错误、和主流教科书《AI: A Modern Approach》在哲学基础和测试理论上的错误长期存在，说明现在的研究和教育体系均存在严重问题。

用人工智能改进教育是好事，但也要认识到人工智能技术的局限性，也要避免误导科学研究、误导社会。当然如果不是这方面的学术精英，说人工智能全知全能的人自己也可能是受害者。

中小学教育的知识很成熟，可以靠“老师录视频，老师编辑知识点，老师给出规划”，在此基础上用人工智能技术提高效率。但光用这种细分知识点的方法，虽能提高应试教育技巧 [20]，却会使学生的思维习惯碎片化，不利于培养系统化和抽象化的科学思维方法。

事实上，从数学、哲学、到科学的可持续发展，需要系统的思想和理论，需要向第一性和完整性发展。目前脑和智力的研究中出现的问题，就是由于教育中忽视了对科学思想方法的训练。而现在不成熟的教育 AI 技术，会使这种问题更加严重。

没有这种思维训练，当然就分析不出 AlphaGo Zero、语言处理和诡辩、和 AI: A Modern Approach 中的问题，也无法从科学哲学的高度抽象概括出亚里斯多德思维模式和计算机思维模式的系统性和结构性问题，甚至会使未来的科学研究误入歧途。

大学主流人工智能教科书 AI: A Modern Approach 的哲学基础和测试理论都有严重错误，AI 本身就有严重问题，怎么能靠 AI 来自我进化、纠正编教科书的大学教授的错误?

思维模式的问题，儒家教育造成的万马齐喑悲剧就是一个典型例子。洋务运动从知识层面作改进，在科学和教育方面的提高也有限，根本不能解决问题。

对高等教育来说，如果靠亚里斯多德的信徒编辑知识点、给出规划，哥白尼、开普勒、伽利略都会被封杀，怎么能发展现代科技?

儒家和亚里斯多德的信徒都是人，至少还具有人的一般智力和变通适应能力。而计算机和人工智能技术，使用过度简化的逻辑，有 Chinese room 等等的问题，不能真正理解语言、哲学、数学、科学等等人类学术的发展和构造。如果让学生从小就服从计算机过度简化的逻辑和流程，人的特质会退化，智力发展会受严重损害，后果更严重。

现在有人工智能教育方面的人士，把对称思维、创造性思维、分类讨论思维、类比思维等等当作思维模式。这些其实只是提供更多思维可能性的思考方法，不包含决定结论对错的 judgement，也不是思维模式。

而思维模式，会决定结论的对错。很多思维模式，如计算机思维模式、亚里斯多德思维模式，会导致错误的结论。所以需要研究很多思维模式的问题，也需要研究科学是怎么避免和纠正很多思维模式的问题。

计算机和人工智能技术，既给教育带来很大帮助和机遇，也带来很大挑战。需要对语言处理、思维模式等等有关人工智能的问题，用科学方法认真研究。在科学验证的基础上，谨慎地设计教育中的人机耦合。

特别需要谨慎的是，中小学教育实际上是红海市场。整体教育规模很难有大幅度增长，前些年在计算机和网络教育上也已经投入很多资金、形成了很多用户 [21]。

而科学新学科的建立和发展，至少可能给高等教育带来局部的增长。所以，人工智能教科书的问题，知识、学术、智力的结构系统理论，如何严格控制条件来设计证伪实验的科学基础，这些研究对未来的发展非常关键。

研究性大学，并不等同于精英大学。科学需要平等讨论和公正实验。平民因此可以靠科学上的努力，弥补一些自己的弱势。而开创科学新领域，探索新的教育和研究方式，更需要多样化的教育，理解不同的教育模式和需要。

十二、不同的教育模式

1989年，新加坡举行了一场辩论赛，辩题是：“儒家思想可不可以抵御西方歪风”。辩论虽然是反方获胜，但评委最后的结论却是辩论双方都没有把问题分析清楚。当然评委自己也无法分析出对错。

辩论赛和AlphaGo系列的围棋比赛一样，虽然有胜负，却不能分辨学术上的对错[22]。辩论赛和Socratic方法，也有本质不同。这作为智力发展的问题，需要进一步研究。

如果把这场辩论赛，和2015年的Bohunt Chinese School实验、及BBC的Are Our Kids Tough Enough?的相关系列电视节目对照起来看，就有意思了。

实验中的中国教师一再提到英国学生落后中国学生三年。这个说法并不准确。有数据说上海学生在2012年PISA评估中最好，领先各国平均学生三年。但2015年和2018年PISA用中国发达的东部两市两省代表中国大陆后，中国大陆发达地区学生的表现，不但落后于新加坡，也落后于台湾等好几个国家和地区，跟上海学生单独得分差距非常大。上海、新加坡的领先，是否跟这两城市的户籍、入境管理和其他因素有关，也需要研究。

所以，用数据作依据，很容易产生误导。这是数据方法需要警惕的问题。Bohunt Chinese School实验，只是经验性的实验，不是科学实验。Bohunt校长本人也谈到实验结果不能说明问题[23]。跟人有关时，特别需要注意经验实验和科学实验的区别。

整体上，新加坡、香港、澳门、台湾、中国大陆的华人聚居区，确实在PISA评估中名列前茅。这里有个文化原因。华人在七世纪开始系统化、制度化的考试，远远领先其他国家，也远远早于十七世纪的科学革命。但华人在考试上的悠久文化，并没有催生出科学革命。

Bohunt实验中的中国教师，把Confucianism（孔儒）当作哲学基础，不但没有说服力，反而在英国学生中造成了很大的困惑，并在一定程度上妨害了他们的教学。本来他们的教学可以更加成功的。

事实上正是孔儒文化直接导致了中国近代在科学上的严重落后。孔儒文化，讲究传承。但跟现代教育的内容，却有严重脱节。这可以用Bohunt实验中的中国教师教授的几门学术课程：数学、科学、英文、中文来作分析。

英文出现和成形，跟孔儒完全没有关系。科学的起源和成形，跟孔儒也基本没有关系。数学的起源和成形，和孔儒也关系甚微。中文倒跟孔儒关系较大。但孔子之前中文的文言文已经比较成熟；以后的发展是百家争鸣，连普通人都有很多贡献；二十世纪初的新文化运动更是通过对孔儒的激烈批判才发展起现代中文。

所以，对英国学生谈文化传承，讲莎士比亚、科学革命、工业革命，才跟课程真正有关系，才能让他们理解在发展中的真实自我。而更深的渊源，需要追溯到毕达哥拉斯、希帕索斯、苏格拉底、欧几里德等等。现在人工智能在哲学基础和测试理论上的问题，以及对智力的结构性和系统性分析，都需要在深度上理解这样的文化传承。

孔儒是主张尊卑有序的。而那几位中国教师反复讲的discipline和competition，其实是比较现代的概念，跟孔儒理念并不相同。事实上正是尊卑观念妨害了他们的教学。

Discipline跟尊卑无关，只跟人在不同场合的角色有关。在Bohunt实验中，很多英国学生对英国教师很尊重，但对中国教师经常嘲弄、甚至愚弄。那些英国教师对学生有管理上的权势；而中国教师来自相对不发达的国家，很容易被认为地位卑下。跟英国学生讲Socratic方法、平等讨论，才能从根本上消除他们的嘲弄、愚弄态度。

那几位中国教师反复讲competition，强调要变得更强。这其实是对达尔文学说的误解。达尔文已经认识到是适者生存，而不是强者生存。科学家和教师，跟权势和财富的拥有者比起来，其实比较弱势，很容易在骨子里受鄙夷。

宋朝灭亡后，郑思肖在《心史》中就提到“一官、二吏、三僧、四道、五医、六工、七猎、八娼、九儒、十丐。”做不了官的教师是儒，在元朝地位比妓女都低。而元朝正是中国历史上最强大的朝代。英国伦敦在2012年刚刚竖起成吉思汗的雕像。

所以，不宜让学生误以为学业优秀就是强者。追求学术上的进步是为了帮助人类更好地适应，是为信仰而服务。社会则应该尊重这种服务。作为信仰和服务，需要认真分辨学术上的对错。教育上是十年树人，更要对学生负责、分清对错，不能颠倒是非。

那几位中国教师，虽然对讲授的专业比较纯熟，但对专业以外的很多问题有些片面、狭隘。对自己的处境、社会中实际存在的尊卑次序、孔儒尊卑有序观念的真

正后果，也未必了解 [24]。

但英国在教育上确实面临问题。十九世纪，英帝国和清帝国，是世界上管理人口最多的两大国家。英帝国虽然战胜了清帝国，但在达尔文之后，其学术研究跟其他发达国家相比，其实在走下坡路 [25]。到了李约瑟、鲁桂珍时代，思想更是有肤浅化的趋势。对不同思维模式及其教育模式，并不真正了解。中国和其他非西方的思维模式和教育模式，并非一无是处，但不是李约瑟和孔儒所讲述的那种优势。Leukotomy 的错误无法纠正，更给未来脑和智力的研究蒙上阴影。

理解不同文化和不同教育模式，需要对智力和文化作结构性、系统性的深度分析，需要新的哲学。现在，美国教育也有类似于英国教育的问题，甚至更严重。Technological Singularity 和主流教科书 AI: A Modern Approach 中的问题，就是这种迹象。

十三、一个有关教育、科技、和文明发展的讨论

科学研究基于平等讨论和公正实验。除了对人工智能设计渐进的、严格控制实验条件的科学实验，本人也设计了一个有关教育、科技、和文明发展的讨论。

牛顿不是科学的创建者，也不是物理学的创建者。牛顿哲学有很大的局限性。马尔萨斯的人口理论、凯恩斯的经济理论、钱学森的农业观点，也极具争议。人口、教育、科技直接影响到智力的发展，影响到文明的未来。学校该如何在重要而极具争议性的理论观点上教育学生，也颇为棘手。

十年树人，百年树木。但误人子弟，也后患无穷。

钱学森的农业观点及其影响，曾广受指责。但在转基因技术发展的新前景下，其观点又变得重要起来，有重新审视的需要。

本人设计了以钱学森的农业观点为导入，并深入展开对马尔萨斯的人口理论、凯恩斯的经济理论、和牛顿的哲学进行讨论，以此来分析和澄清有关教育、科技、文明等有关人类未来的许多重要问题的研究。

本人邀请袁隆平教授、中国工程院、美国科学院、和联合国粮农组织，在转基因技术发展的新前景下，对钱学森的《展望十年——农业发展纲要实现以后》和《可以实现的理想》两篇文章发表新的分析和观点。

一旦收到他们的答复，本人会公布自己的分析评论。“不作评论”也属于有效答复。

本人不是农业专家。本人只对科学的方法论和智力的结构性、完整性等等相关问题进行分析研究。

十四、结语

科学哲学在 Karl Popper 以后，基本停滞不前。导致学术界错误泛滥，也使脑和智力研究在哲学基础及测试理论和方法上都误入歧途。

本文引入对人类知识和学术的结构性和系统性分析、和第一性和完整性的复杂性分析，作为脑和智力研究的科学哲学新基础。以此分析学术上的种种错误，设计严谨的科学实验对人工智能技术证伪。

学术界流行的很多观点，破坏了学术语义和科学实验的严谨。如果不纠正，会使诡辩泛滥，在教育、学术研究、医疗、司法等等许多方面，引发全局性的危机。

BRAIN Initiative 项目，无法对智力的完整性作系统性、结构性、和复杂性的分析，所以无法分析人格、多重人格、思维模式，无法分析很多重要的心理问题、及其引发的精神疾病，也无法为脑和智力研究提供整体的科学基础。

把智力研究转化成科学的关键，在于形成准确描述智力的语言概念、哲学基础、数学形式、和证伪实验。经验测试不是科学实验，科学结论必须接受别人严格控制条件的科学实验的证伪。拒绝别人的质疑和证伪实验，也是在拒绝进一步发展的机会。

为智力研究建立结构性、系统性的科学分析，和远离测量极限、严格控制条件的实验基础，可以为科学研究带来新的生机和动力，是继数学、物理科学、生命科学以后的又一重要研究方向。这也关系到正在出现的人工智能挑战。挑战既包括人工智能在某些领域超越人类，也包括滥用不成熟的人工智能可能带来的严重后果。

智力的基础科学研究，不但可以帮助人类和人工智能更好地配合、发挥出各自的优势，而且可以带动脑科学、生物学、语言学、心理学、哲学、数学等等一大批学科的发展，为多样化的教育带来新的机会。

智力研究的重要性是其他研究无法替代的。不能因为其他研究有些进展，就轻视、排斥智力的科学基础研究。如果自己不懂，也不让别人研究，只会导致血腥的红海竞争。

在 The Development of Liberal Arts and Sciences 一文中，本人将分析蒙昧主义和诡辩的土壤，科学对人文的关系和依赖，以及艺术对思维模式形成、发展、和转变的作用。

~~~~~~~~~~~~~~~~~~

[1] 人工智能的主流教科书 AI: A Modern Approach 的第三版，从哲学基础中删掉了对苏格拉底和柏拉图的介绍，把亚里斯多德思维模式当作正确的思维模式: “Aristotle… was the first to formulate a precise set of laws governing the rational part of the mind.”(On page 5)。

由于这个思维模式的错误，该书提倡的 Wind Tunnel approach 也只是经验方法，不能科学地测试人工智能技术的可靠性和安全性。

该教科书在哲学基础和测试理论上的错误，实际上也覆盖了整个人工智能行业。

[2] 事实上太阳系的重心就在太阳附近。以太阳为基点的参照系比以地球为基点的参照系更接近惯性参考系。如果采用地心说或第谷的理论，就无法发展出牛顿的物理体系，无法渐进地发现科学规律。

第谷的观察工作仍有价值是侥幸，因为地心说、日心说、第谷理论在观察方法上没有争议。哥白尼、第谷、开普勒时期的天文学，还不是严格意义上的科学，所使用的方法也不是严格意义上的科学方法。

而很多科学实验，需要严格控制实验条件。实验结论要根据实验条件来推导。错误的实验条件会导致实验无效或错误的结论。

[3] 按政府报告 Leucotomy in England and Wales, 1942-1954，该手术的 9284 名患者中，41%恢复或有明显改善，28%稍微改善，25%显示无变化，2%恶化，4%死亡。

但这个统计很容易被误解。如此破坏性的脑手术，当然会破坏 personality 和 intelligence。所以，这里的 25%显示无变化，不是指 personality 和 intelligence 未受破坏，仅指为控制精神状态而特殊定义的 clinical condition 无变化。同理，41%恢复或有明显改善和 28%稍微改善，也只是特殊定义的 clinical condition 有改善，personality 和 intelligence 实际上都有损伤。

本来，leucotomy 手术就是为了控制精神状态，而不惜损伤脑功能。

也有冠名 Renato M.E. Sabbatini 教授的文章指: Even lobotomy's preponents admitted that only one third of the operated patients would improve, while one-third remained the same, and one-third got worst 。这些数据来源不明，跟 Leucotomy in England and Wales, 1942-1954 的差距也很大。（http://www.cerebromente.org.br/n02/historia/lobotomy.htm）

但这里的 one third would improve 和 one-third remained the same，也仅仅指特殊定义下的 clinical condition，不是指 personality 和 intelligence 未受破坏。

所以，由于缺乏正确描述 personality 和 intelligence 的语言，就无法对 leucotomy 手术和使用基因技术的 BRAIN Initiative 方法的破坏性后果作研究。这种情况急需改变。

[4] 对学术论文的审核，只代表审核者的观点，不是学术上的定论。所以，学术上的苏格拉底方法很重要。要给被审核者和其他人反驳的机会。
~~~~~~~~~~~~~~~~~~

对科学论文的审核，也跟数学论文不同。数学论文只能靠 peer review。科学论文中的结论，可以用新的证伪实验来推翻。其错误，从论文本身可能看不出。所以，对科学论文的 peer review，并不能保证其正确性。要让别人有机会设计新实验来证伪。

对科学论文的审核，可以发现推理问题，实验数据是否真实等等。但推理问题只能引起注意，有可能是出于理解有误。所以下结论要经过对话、反驳、和实验证伪。AlphaGo Zero 论文中的 superhuman 宣称，是针对 generic 的 human，并没有相应的证据。本人却可以设计严格控制条件的实验来证伪此宣称。但 AlphaGo Zero 方面至今拒绝对话和实验。

[5] 我选择围棋作实验工具，是因为可以用围棋验证人工智能和计算机的一些基础科学理论，跟军事战略和战争艺术无关。围棋是和平的竞技，高度抽象而又直观。既古老又现代，能吸引亚洲以外的电脑研究。已经成为跨文化研究、交流的平台。

第一轮实验后，开发者可能修改软件。所以有些实验证据只能在第一轮实验中收集。第一轮实验需要跟最强版本电脑围棋作较大规模的实验，以收集足够的实验证据。

我对 AlphaGo 系列的实验不是白盒测试，不涉及知识产权。如果 Google 愿意公正地检验，可以仿照美国国会图书馆先例，先向国会图书馆递交 AlphaGo 柯洁版、AlphaGo Zero、和最新 AlphaGo 版本所需要的所有硬件、软件和文档，能捐款给图书馆用于管理和维护则更好。以后在第三方场地实验时，可以向国会图书馆核对软件、硬件和文档等版本。Human level artificial intelligence 的检验，是人类发展史上的里程碑。如果无人能对 AlphaGo 系列证伪，国会图书馆就有继续保留其版本的价值。其他人有实验想法，可以先做。如果其他人都失败了，就没有道理拒绝我的实验方案。

这也为美国政府了解人工智能的最新进展提供了渠道，并为以后的人工智能测试建立了一个范式。一旦相关实验设施建立好并熟悉了这种流程，以后重要的人工智能测试费用会大大降低。

[6] 因此实验涉及到某些敏感因素。目前我设计的实验方案，只能发现一部分问题，不能发现全部问题。所以需要进一步的研究。而我的实验，可以提供进一步研究所需要的证据。所以要防止实验前系统被针对性地改动，破坏了进一步研究所需要的证据。

科学要求通过不断证伪来验证、改进结论。而 Demis Hassabis 公开宣布 AlphaGo Zero 不再参加比赛，拒绝了进一步的验证。事实上 AlphaGo Zero 从来没有严格控制条件下的实验证据。所以如果 Deepmind 拒绝验证，会破坏科学的方法和基础，损害学术道德。

再考虑到 Deepmind 的 ethics board 仅一个星期就被解散，Deepmind 在道德上言行是否一致有很大疑问。我对证据被破坏的担心并不多余。所以，Google 向国会图书馆递交 AlphaGo 系列的软、硬件和文档后，我才能递交我的实验方案。

[7] 目前能见到的 AlphaGo Zero 棋谱中，对 AlphaGo Master 的二十局是 AlphaGo Zero 最后版（或稳定版），有很不寻常现象：AlphaGo Master 执黑的十盘棋中，布局至 16 手完全一样。而 AlphaGo Zero 执黑的十盘棋中，布局至 18 手完全一样，AlphaGo Zero 输棋的 14、16 局，至 45 手完全一样。所以有如下问题：

1）在 Nature Magazime 上发表两篇 AlphaGo 系列文章后，Deepmind 有没有在这二十局棋里把 AlphaGo Zero 和 AlphaGo Master 对局的情况，全面准确地告诉公众？如果答案是，则：

2) 公布的这二十局棋谱中，所出现的布局高度类同，是由于 AlphaGo Zero 接近围棋最优解，还是陷入 local trap？如果认为是前者，需要用严格控制条件的严谨实验来验证。

[8] 如在 The Guadian 访谈中他谈到 a meta-solution to any problem：

"Demis Hassabis ... is deadly serious when he tells me he is on a mission to 'solve intelligence, and then use that to solve everything else'.

……

'One way of thinking of AGI is as a process that will automatically convert unstructured information into actionable knowledge. What we're working on is potentially a meta-solution to any problem.'",

from https://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago。

[9] 许峰雄在 Cracking Go 文中更早提出。但他的依据很多来自开发 Deep Blue 的经验，对围棋不一定成立。我的理论分析，不依赖经验，反而能成立。我的理论分析还可以进一步延伸：AlphaGo 系列虽然赢棋，但只要不是最优解，就可能走的是无理棋；可以设计严格控制条件的严谨实验来检验 AlphaGo 系列是否有通用智力潜力，是否超人。

[10] 由三体问题的研究发现，行星的真正轨道极为复杂，传统数学都难以完整表达或逼近。所以，下文提到：Universal approximation theorem 的名称，也很容易误导；而可以抽象成 Turing Machine 及其各种变种的计算机技术，也有本质的局限性。

[11] In God We Trust 是美国的格言。这里提到的 God 是全能神。现代科学可以构造多个具有哲学完整性且无法证伪的科学理论。这些无法证伪的科学理论，使全能神的存在，也无法被证伪。但人不是全能神，人无法完整理解全能神，所以说 Trust。

数学是人的一种智力理论，不是全能神。数学的局限性，早已被证明。所以，In Math We Trust 并不正确，是迷信。In Math We Trust 在精英专业研究人士中的流传，暴露了科学哲学教育的缺乏和现代教育的问题。

[12] 登月项目是工程。需要定义明确的实施目标。而脑研究是基础科学研究，需要通过科学实验的证伪来不断修改假设。所以，工程和基础科学研究的方法完全不同。出现这样的混淆，是教育和研究严重失败所造成的无知。

[13] 在政治宣传、商业宣传中，诡辩和语义的混淆比比皆是，难以清除。所以，科学研究和司法，需要跟政治宣传和商业宣传严格区分，保持应有的独立性。

[14] 数学上的欧几里德几何和非欧几何，是作为有效理论并存的。数学可以作为科学的工具，但本身并不是科学。事实上，非欧几何的出现，就明确宣告欧几里德几何不是科学，而仅仅是数学，数学也不具有哲学完整性。哥德尔定理更说明数学不能判断科学上的正确性。而科学上的多种替代性理论中只有一种是正确的，所以科学有完整性。

[15] 所以，达尔文的理论，按其性质的不同，可以分成两部分：

一部分是演化的规律，这是科学理论，可以用实验检验和证伪。虽然充分地检验，需要很长时间。这部分争议较少。所揭示的生物本质和规律，奠定了区别于博物学的生物科学基础。

另一部分是物种的起源。这和宇宙大爆炸理论一样，实际上是对历史的逆向推导，是历史研究。虽然可以用到很多科学方法，但历史研究本质上不是科学，只能提供某种程度的证据，不能用严格控制条件的实验验证或证伪。当然，不是科学并不等于就是错的，只是说这种理论达不到科学理论的标准。但历史研究不能用实验验证或证伪，确实也造成对一些年代较远的事情的结论经常出现变化甚至反复。

[16] 亚里斯多德基于特征的生物分类法，是经验性的主观方法。这种主观方法会掩盖生物分类法中的歧义。现代生物分类法试图引入基因等客观因素，却暴露了生物分类法中的歧义是本质性的，不但无法用逻辑来消除，其歧义甚至扩散到顶层类别。所以，生物分类法只是语言层次的经验知识，无法达到哲学完整性。不是严谨的科学，也不是科学哲学。

[17] 高斯不但没有公布他自己的非欧几何想法，也没有公布 1819 年就交给他的 Ferdinand Schweikart 的类似想法。所以高斯对非欧几何，不但无功，反而有过。原因就在于高斯太注重坚实的技巧性，不了解尚未成熟的思想体系的价值，也不重视苏格拉底方法对学术发展的作用。高斯不但缺乏学术眼光，也达不到罗巴切夫斯基的思想境界和层次。

罗巴切夫斯基和鲍耶推翻欧几里德几何的唯一性，已是 1830 年前后了；他们在非欧几何上贡献最大。黎曼几何在物理相对论研究中，有价值。但黎曼在非欧几何中，只是作延伸性工作，贡献次于罗巴切夫斯基和鲍耶。贝尔特拉米为非欧几何带来新的语义和完整性，也很重要。

由于被高斯埋没，Ferdinand Schweikart 可能没有对非欧几何的发展产生多少实际影响。我也不知道如何评价他的工作。

[18] 《九章算术·方程》:“正负术曰: 同名相除，异名相益，正无入负之，负无入正之。其异名相除，同名相益，正无入正之，负无入负之。”

也有观点认为，《九章算术》之前，中国已经出现负数概念。

[19] 中国古代百家争鸣时期，就有阴阳家，《道德经》和解释《易经》的《十翼》也都论及阴阳。阴阳概念的真正起源，有待考证。但《易经》体系的阴阳并不导出负数，而是: 老阳数九，老阴数六，等等。

阴阳概念，不但可以很容易导出正负数，也为现代正反物质的理论提供了一个很自然的通俗理解语言和思维模式。但能相互抵消的正负数，是在阴阳概念基础上的飞跃。现代正反量子湮没的理论，又是对可抵消的正负数概念的飞跃，因为量子湮没是伴随有能量形态的转化的。从中正好可以看到从玄学、到数学、到科学的思维模式的演变和发展。

但阴阳和玄学不是科学哲学语言，不是具有完整性的严格意义上的哲学。最多只能算语言层次的准哲学或哲学性思想，进一步发展的潜力有限。在形成正负数以后，中国古代数学也没有在这个方向有多少后继发展。所以，即使准哲学或哲学性思想能帮助形成一些新的学术思想，准哲学或哲学性思想本身也需要不断发展，需要新的语言概念。

中国古代没有像希帕索斯那样对无理数的准确发现，对类似无理数的概念没多少抵触，但也不理解其价值。这里有哲学和思维模式的不同，受此影响智力的发展轨迹也不同。

张首晟教授生前提到过阴阳、正负、和量子湮没。在此注明一下，作为对他的纪念。本人很多年没有参加投资创业活动，以前也没有看到张教授的这个观点。他去世后搜寻他的信息，才发现他有类似观点。也许科学的追寻者本来就可以心灵相通。写上述内容，是为了整理中国古代学术思想。写作时仅通过 wikipedia 查证正负数概念出自中国。

不过本人的观点跟张教授也有所不同。本人着重指出从不导出负数的阴阳、到相互抵消的正负数、到伴随能量转化的量子湮没，是从玄学、到数学、到科学的不断发展，其哲学性基础有质的变化。这里包含着思维模式演变的临界点。而张教授等一些学者认为准粒子和基本粒子没有本质区别的观点，是物理学界的争论，本人不发表看法。

准粒子跟 emergent phenomena 有关，是凝聚态物理中的重要概念。在物理学中，数学推导不能作为科学结论。但物理学中有海森堡测不准原理。如果临近测量的极限，会有很多争议，有些甚至无法解决。但离测量极限尚有一定距离时，仍有可能通过精化测量的尺度来区分。

[20] 奥林匹克数学竞赛和朗道势垒也是种应试，只是难度较高。奥林匹克数学竞赛在科学性和理论深度、广度、高度上也远远不及朗道势垒。只是一味追求难度，反而会使教育丢失更重要的目标。

最近用计算机和人工智能技术改进中小学教育的努力，重新重视了教育的基本面、和对基础知识的解析和积累。但计算机和人工智能技术有局限性，也需要深入研究。

一个重要的问题是诡辩。近年诡辩再次流行，甚至在学术界对“白马非马”是诡辩都出现混淆。诡辩入侵教育，会给智力发展、学术研究、和司法等等造成严重危害。

[21] 无人汽车的目标市场，其实也是红海市场，因为整体汽车市场的规模，很难再有几十倍、几百倍的增长。这跟计算机市场的早期发展，截然不同。

计算机技术从 mainframe、到 personal computer、到 smartphone，都是在开拓新的市场。但现在 smartphone 已趋于饱和，技术上的更新换代，已难使整体市场大幅度增长。

所以，虽然容易理解 IT 业界急切寻找新增长点的动机，也容易理解很多产业在红海竞争下寻找自动化来降低成本的压力。但人工智能的主要目标市场，并不是像计算机产业早期那样的新兴市场。

栗浩洋在分享人工智能市场经验时，也强调“颠覆现有的成熟商业模式”、“不要完全去造一个全新的商业模式”。但这是红海经验，幸存者会很少。

他的另两条经验是“容错率高，走中间路线的公司”、“有限可能性的领域”。也就是说，他知道目前的人工智能技术有很大的局限性。

应试教育的容错率高，因为绝大部分学生的目标并不是考 100 分，一般考试 100 分中得 60 分就能通过。但教育本身的容错率并不高，特别是思维模式的训练形成和诡辩的识别，影响到社会的兴衰。只是十年树人，其后果显露需要很长时间。

工业自动化，因为应用于产品，成败后果主要由企业和员工自己承担，对社会的危险比较小。实际上也一直在渐进发展。而教育、医疗、司法的服务对象就是社会上的人，无人汽车跟社会上的人的互动也非常密切，需要特别谨慎。

目前的人工智能教育系统是作为对正规教育的补充，其问题需要靠正规教育或其它方式的教育来弥补和纠正。长期来看，人工智能对教育的影响、冲击、挑战不可避免。

所以，对人工智能无法避免的局限性、以及人机耦合的可能模式，需要在学术上认真、深入、系统地作前瞻研究。忽视、排斥这样的研究，人的教育出了严重问题，就晚了。

本人只从事智力和教育的研究，无意参与商业竞争。本人赞成用计算机和人工智能技术对教育变革作探索。指出这些问题，是为了更好地探索、变革、和发展。

[22] 辩论的局限性，在中国古代的百家争鸣时期就已经暴露出来。战国时期倪说、公孙龙以“白马非马”之说辩赢其他学派人士，包括当时重要学府稷下学宫的学者。

但不加限制条件的“白马非马”会产生误导。用不加限制条件的“白马非马”来辩赢别人，更是诡辩。而其他各派学者却无法用学术分析反驳，说明辩论方法的局限性。

《韩非子》的作者也无法用学术分析来反驳“白马非马”，只是说可以用法家式立法来压制“白马非马”。但学术问题应该用学术方法解决。“白马非马”在某些抽象限制条件下是成立的，这就是其他战国学者无法反驳“白马非马”的真正原因。光用法家式立法来压制“白马非马”，抽象思维、逻辑、和语言学术就无法继续发展，更无法发展出欧几里德几何这样的数学体系、以及需要更高抽象程度的科学方法。

所以，百家争鸣虽然促成了中国古代的学术繁荣，但并不能解决一些重要的学术问题。光有百家争鸣是不够的。进一步分辨是非对错，需要有完整性的哲学和科学方法。

要反驳各种诡辩的错误，需要对复杂的抽象限制条件作仔细的分析。而计算机技术在处理抽象和高阶逻辑上有严重问题，所以无法分析处理复杂的诡辩。

在计算机和网络时代，“白马非马”不是诡辩的说法，出现在一些主要的网站上。这其实说明了计算机思维模式的问题，因为计算机使用的是过度简化的逻辑。由于“白马”和“马”是两个不同的词和概念，熟悉计算机编程的人很容易简单化地认为“白马非马”是对的，就接受了；而对“白马非马”在语言常识中其实是错的、用不加限制条件的“白马非马”来辩赢别人是诡辩，反而看不到了。

受计算机思维模式及其过度简化逻辑影响的人，反而难以分辨“白马非马”这种明显的诡辩。所以，未来教育和研究，也需要分辨计算机思维模式的问题。否则诡辩盛行，司法、教育、研究、医疗等等方面，都会受严重损害。

[23] “尼尔·斯特劳格尔校长说，在去年的这部纪录片里，中英教育的差异被人为放大。他说，参与拍摄的英国实验班学生的真实学业水平，本就比一般的其他班学生强，这是 BBC 特意提出的要求。”（http://shanghai.xinmin.cn/xmsq/2016/04/18/29861595.html）

即使实验班学生原来学业水平跟其他班学生相仿，用一次考试结果给教育下结论，也太片面。这样的经验实验，主要作用是让人看到问题、引发思考，而不是以此下结论。

[24] 中国民间本来就有“笑贫不笑娼”的说法，跟孔儒的尊卑次序并不相同。司马迁的《史记·货殖列传》中还有“夫用贫求富，农不如工，工不如商，刺绣文不如倚市门”。所以孔儒尊卑有序的观念会有什么真正后果，可想而知。

像开普勒那样死于贫困的学者反而不受推崇，还受封杀。所以，社会对服务于科学的研究者不但有尊重和资助的责任，还需要遵循平等讨论、公正实验的严格科学方法，才能给他们机会来发展科学。

现在美国面临的教育改革，教师工会是很大的争议，却至今没有系统的好方法来解决问题。没有教师工会、社会、和政府的干预和来自社会、政府的资助，教师的工资可能会很低。而没有社会和政府资助，研究基础科学的科学家处境甚至可能比教师更糟，根本没法自食其力。开普勒不得不靠天文学的“应用”占星术来乞讨。

所以，没有社会或政府干预和资助，教师和基础科学家实际上会沦落到“九儒”、“十丐”的地步。这是李约瑟没有理解清楚的方面之一。所以说他肤浅。

[25] 由于英帝国的制度和社会问题，联合王国本土、加拿大、澳大利亚、新西兰以外的众多非欧洲裔英帝国人口，对英帝国的智力和科学发展，实际贡献很小。所以，英帝国的核心人口，规模其实不大。这也是衰退出现的重要原因。

而现在，中国和印度基本普及了教育。在人口规模上的优势也渐渐体现出来。中国因为二十世纪的新文化运动和思想解放运动，在哲学和科学上有了较多的独立思考，在最近几十年中展现出强大的发展势头。印度的未来也值得关注。